



HAL
open science

Editer l'Encyclopédie au 21e siècle : un projet d'édition numérique, critique et collaborative

Alexandre Guilbaud, Irène Passeron, Vincent Barrellon, Olivier Ferret

► To cite this version:

Alexandre Guilbaud, Irène Passeron, Vincent Barrellon, Olivier Ferret. Editer l'Encyclopédie au 21e siècle : un projet d'édition numérique, critique et collaborative : Bilan et perspective de la recherche dix-huitiémiste. *Dix-Huitième Siècle*, 2014, Des recherches dix-huitiémistes aujourd'hui, 46, pp.153-166. hal-01258324

HAL Id: hal-01258324

<https://hal.science/hal-01258324>

Submitted on 1 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉDITER L'ENCYCLOPÉDIE AU 21^e SIÈCLE : UN PROJET D'ÉDITION NUMÉRIQUE CRITI- QUE ET COLLABORATIVE

Si les premières tentatives, en France, de traitement informatique des textes remontent aux années 1960, c'est surtout dans les années 1990-2000, avec l'évolution rapide d'internet et l'augmentation des capacités de stockage des données, que sont véritablement nées les Humanités numériques (ou *digital humanities*) et, en particulier, qu'un intérêt croissant s'est manifesté pour les possibilités offertes par l'outil électronique dans le domaine de l'édition¹. Il s'agit là d'un axe de recherche actuellement en plein essor, comme en témoignent non seulement des entreprises de *numérisation* effectuées à grande échelle (Gallica, Google Books), mais aussi, plus récemment, des entreprises éditoriales *numériques* adossées à des projets scientifiques.

Malgré leur développement incontestable, la question du statut des humanités numériques ne fait cependant pas consensus aujourd'hui² : elles sont, à une extrémité de l'échelle, considérées comme le prolongement naturel des sciences humaines et sociales, pour lesquelles l'utilisation d'outils informatiques deviendrait progressivement indispensable ; à l'autre extrémité, le « Manifeste des humanités numériques », signé par plus de 200 chercheurs, intellectuels et représentants d'institutions en 2010, revendique pour les *digital humanities* le statut de « transdiscipline ». C'est autour de cette seconde perspective, qui implique fondamentalement l'interrogation mutuelle de deux champs disciplinaires (sciences huma-

1. Lou Burnard, « Du *Literary and linguistic computing* aux *Digital Humanities* : retour sur 40 ans de relations entre sciences humaines et informatique », in *L'édition électronique ouverte*, Marseille, Cléo, 2009.

2. Benjamin Caraco, « Les digital humanities et les bibliothèques : un partenariat naturel », *Bulletin des bibliothèques de France*, t. 57, n° 2, janvier 2012, p. 69-73.

nes et sociales et informatique) en vue d'un objectif commun (une édition numérique par exemple), que s'articulent actuellement les débats les plus féconds, notamment sur le renouvellement des méthodes et des pratiques pour l'enseignement et la recherche³. Pour expliquer rapidement ces enjeux, prenons l'exemple, central et peut-être même paradigmatique, de l'édition critique (critique au sens large, textuel et explicatif) de textes anciens.

Il s'agit désormais d'inventer de nouvelles formes d'édition électroniques susceptibles d'accompagner et de susciter de nouveaux types de lecture⁴, et, pourrions-nous ajouter d'écriture, dans la mesure où l'un des enjeux, pour les éditions critiques numériques est aussi de se constituer en outils de travail et de recherche – ce qui implique par exemple l'intégration, dans les interfaces d'édition, d'outils d'appropriation des contenus (annotation, classement, écriture de commentaires libres, etc.) à destination du lecteur et de l'éditeur⁵.

L'édition numérique est donc aujourd'hui confrontée à un double enjeu lié à sa double nature : d'une part, éditer un ouvrage ou un ensemble d'ouvrages selon des codes hérités de l'édition papier (point de vue statique) ; d'autre part, éditer un texte en utilisant les ressources numériques, en particulier, d'interrogation et de navigation (point de vue dynamique). Le premier enjeu consiste à faciliter et guider une lecture suivant un ordre linéaire, lié à la

3. Voir la définition du *MA/MSc in Digital Humanities* proposé par l'University College of London : <http://www.ucl.ac.uk/dh/courses/mamsc>, ou celle proposée dans le « Manifeste des Humanités numériques », THATCamp 2010, accessible en ligne à l'adresse <http://tcp.hypotheses.org/318> (consulté le 1^{er} juillet 2013). Dans une perspective de définition d'un paradigme nouveau plutôt que d'extension d'un classique, voir également la notion « Humanisme numérique » (Milad Douehi, *Pour un humanisme numérique*, Seuil, Paris, 2011).

4. Dans le sillage des réflexions autour du renouvellement de la philologie, marquée par l'élaboration d'une linguistique de corpus désignée sous l'appellation de *philologie numérique* (voir François Rastier, *Arts et sciences du texte*, Paris, PUF, 2001, p. 73-97).

5. Voir Simon et al. « The YUMA Media Annotation Framework », *Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science*, Volume 6966, 2011, p. 434-437 ; Audenaert et al., « CritSpace: A Workspace for Critical Engagement within Cultural Heritage Digital Libraries », *Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science*, vol. 6273, 2010, p. 307-314.

matérialité et aux traditions du support papier. Ce qui est donné à lire se décline suivant des choix et des principes (de texte de base, de transcription, de présentation) – et l'on sait qu'il est illusoire, voire trompeur, de faire l'économie de leur explicitation. Le second enjeu est de permettre une lecture suivant un ordre virtuel, multiple, dont la multiplicité est liée aux possibilités d'interrogation du texte, elles-mêmes liées aux métadonnées éditoriales qui structurent le texte et aux outils informatiques qui les interrogent.

Le changement de support implique dès lors une réflexion profonde sur le traitement des textes, sous-tendue par une interrogation sur ce qu'est un texte, considéré comme un objet culturel dont il faut non seulement respecter l'intégrité, mais aussi restituer l'inscription dans un contexte susceptible d'en permettre l'interprétation : sont ici en jeu les questions de la textualité et de l'intertextualité. L'établissement du texte, selon les démarches scientifiques de l'édition critique, implique, au-delà de l'exactitude, entendue au sens de fidélité à l'original, la définition des modalités de sa transposition la plus adéquate possible, minimisant la déperdition d'information, entre son support matériel d'origine et le support numérique. Il repose également sur l'élaboration d'une procédure d'encodage⁶ approprié qui rende compte de son caractère polysémique intrinsèque : un texte n'est pas un simple ensemble de données (*data*), mais un assemblage de données *structurées* qui font système entre elles⁷. L'encodage est enfin intimement lié à la conception des logiciels qui doivent, en les exploitant, offrir des parcours de lecture, d'interprétation, ainsi que des outils de recherche et d'analyse adaptés au corpus édité et capables, à terme, de faire émerger de nouvelles perspectives de recherche.

Un dernier enjeu concerne quant à lui la nécessaire évolution des méthodes et pratiques de travail de l'éditeur critique, c'est-à-dire du chercheur en Sciences humaines et sociales, dans ce nouveau contexte numérique. Il s'agit, dans ce cadre, de juguler l'écart susceptible de se creuser entre, d'un côté, la démarche et les

6. Ce codage doit tenir compte de l'existence d'un standard international, la TEI au format XML, préconisé pour ses bénéfices en termes de pérennisation et d'interopérabilité des données.

7. F. Rastier, « Que cachent les "données textuelles" ? », *Neuvièmes Journées internationales d'analyse statistique des données textuelles* (JADT), 2008, p. 13-26.

compétences d'un chercheur spécialiste d'un corpus particulier et, de l'autre, les technologiques numériques déployées par les ingénieurs et les informaticiens pour permettre l'édition électronique de ce corpus. Ici encore, un étroit dialogue est donc requis entre Sciences humaines et sociales et informatique, de façon à déboucher sur la conception de mécanismes (et donc d'interfaces) de saisie (et donc d'édition) non seulement directement et facilement utilisables par les chercheurs, mais aussi repensés en fonction de l'émergence de nouvelles solutions informatiques, telles que les plateformes de travail collaboratives.

Une telle ambition suppose une nouvelle méthodologie de travail dont la rigueur et la pertinence éditoriales et scientifiques repose nécessairement sur l'établissement d'un dialogue étroit, au sein d'une équipe résolument pluridisciplinaire, entre chercheurs en Sciences humaines et sociales, de chercheurs en Informatique et d'ingénieurs. Le projet ENCCRE (Édition Numérique CRitique et Collaborative de l'*Encyclopédie*) nous permettra ici d'en illustrer les modalités sur plusieurs cas concrets. Il se fonde, pour ce faire, sur une politique éditoriale dont il convient préalablement de rappeler les principales exigences. Il s'agit ici de réaliser une édition numérique de l'*Encyclopédie* :

- complète, en libre accès, qui prenne en compte la matérialité de l'ouvrage (donnant accès à la constitution réelle de l'ouvrage publié, par feuilletage du début à la fin ou par ouverture au hasard⁸ de la reproduction photographique), l'ensemble de ses éléments constitutifs (depuis la page de garde jusqu'aux *errata* et privilèges, les textes et illustrations, dans les différentes versions d'une édition donnée) et le réseau de liens internes à l'ouvrage ;
- critique, c'est-à-dire : articulée étroitement aux recherches passées et aux travaux en cours sur l'*Encyclopédie* ; « fidèle », car la séparation nette entre données encyclopédiques et données critiques commande la façon dont les données sont structurées et codées, puis dont elles sont saisies, affichées et

8. C'est le cas par exemple des ouvrages accessibles sur archives.org, et dans une moindre mesure, mais mieux référencés, sur Gallica.

- rendues interrogeables ; et enfin justifiée (par des notices et notes signées et datées de leur rédacteur) ;
- ouverte, c'est-à-dire adaptée à plusieurs types de lecteurs, à l'enseignement et à la recherche, en proposant plusieurs niveaux de lecture et modes de parcours de l'œuvre ;
 - collaborative, par le biais d'une interface adaptée et fournissant des outils de recherche⁹.

On commencera par souligner le renouveau international des études sur l'*Encyclopédie* attesté depuis deux ou trois décennies et dont témoignent nombre d'études florissantes actuellement : qu'il s'agisse des revues, comme les *Recherches sur Diderot et l'Encyclopédie*, mais aussi comme le *Recueil d'études sur l'Encyclopédie et les Lumières*, des colloques, des séminaires réguliers et suivis, des thèses en cours.

Partons ensuite d'un constat simple : il n'existe pas d'édition critique de l'*Encyclopédie*, éditée par Diderot et d'Alembert, et dont les dix-sept volumes de textes et onze volumes de planches in-folio qui la constituent paraissent entre 1751 et 1772. Ni d'édition papier¹⁰, ni encore d'édition numérique. L'*Encyclopédie* a marqué son siècle et les suivants, mais les innovations qui ont fait sa puissance constituent autant de défis à l'édition numérique : c'est une gigantesque entreprise collective « par une société de gens de lettres », et le nombre de savants contemporains qui y collaborent est l'indice même de sa modernité ; elle se présente sous la

9. Diderot évoquant la nécessaire collaboration des savants « spécialistes » à l'*Encyclopédie* écrivait : « Quand on en vient à considérer la matière immense d'une encyclopédie, la seule chose qu'on aperçoive distinctement, c'est qu'elle ne peut être l'ouvrage d'un seul homme ». Et, plus loin : « qui est-ce qui définira exactement le mot conjugué, si ce n'est un géomètre ? le mot conjugaison si ce n'est un grammairien ? le mot azimuth si ce n'est un astronome ? le mot épopée si ce n'est un littérateur ? » etc. (Diderot, *Ceuvres complètes* DPV, t. VII, p. 175-176). De même, dans le projet ENCCRE, l'historien des mathématiques annotera les articles de mathématiques, l'historien de la grammaire les articles de grammaire, etc.

10. La Table du Pasteur Mouchon, publiée en 1780 constituent peut-être la seule forme d'édition critique existante permettant de naviguer dans les plus de 70.000 entrées du dictionnaire (voir Pierre Crépel, « Peut-on enfin brûler le pasteur Mouchon ? », *Recherches sur Diderot et sur l'Encyclopédie*, 31-32, p. 201-232).

forme d'un dictionnaire ; les arts mécaniques y étant à l'honneur, les planches sont l'un de ses atouts ; à l'image de la *Cyclopaedia* de Chambers, dont elle était censée, à l'origine, être la simple traduction, les nombreux renvois doivent permettre la circulation à l'intérieur de l'ensemble ; elle se veut un recueil critique, c'est-à-dire une compilation raisonnée. En effet, dans la lignée de la floraison des dictionnaires universels¹¹, elle entend donner à ses lecteurs le meilleur des définitions et descriptions existantes, non une simple accumulation de références érudites.

Une édition critique doit donc rendre compte, entre autres, de la multiplicité des auteurs, de leurs différents statuts, de l'incertitude qui pèse sur l'attribution de tel ou tel passage à l'un de ces auteurs, de la justification des attributions. Elle doit bien sûr donner à lire ce qui était, matériellement et de façon manifeste, présent sous les yeux du lecteur des volumes in-folio : la date d'édition, les textes préliminaires, les variations typographiques. Grâce aux facilités offertes par le support numérique, elle doit rendre possible une lecture simultanée de l'original et de la transcription, un accès plus rapide aux articles et aux planches auxquelles le texte renvoie et, surtout, permettre de relier entre eux les résultats des nombreuses recherches dont l'ouvrage a été l'objet¹², et sur cette base, d'en développer de nouvelles.

Pour atteindre cet objectif, l'équipe internationale du projet ENCCRE, associant ici plusieurs chercheurs et un ingénieur humanités numériques, s'est attelée à un minutieux travail consistant à dresser la liste des éléments qui doivent être décrits, à définir les critères selon lesquels cette description doit être effectuée, et à préciser comment ces éléments sont reliés les uns avec les autres. Simple en apparence, cette étape de structuration des données n'en est pas moins tributaire de toutes les études déjà entreprises sur l'*Encyclopédie*.

11. Voir Marie Leca-Tsiomis, *Écrire l'Encyclopédie*, SVEC, Voltaire Foundation, Oxford, 1999, rééd. 2005.

12. Notamment R. N. Schwab, W. E. Rex, J. Lough, *Inventory of Diderot's Encyclopédie*, Voltaire Foundation, Oxford, 1971-1972, 6 vol ; John Lough, *Essays on the Encyclopédie of Diderot and D'Alembert*, London, Oxford U. Press, 1968 ; Jacques Proust, *Diderot et l'Encyclopédie*, 1962, dernière édition Albin Michel, 1995 ; etc.

Soit un lecteur naïf, entendons par là, qui accède à l'édition numérique par un des innombrables chemins du Web sans autre connaissance de l'ouvrage, et soit l'article suivant :

DECENNA ou DÉCURIE, (*Hist. anc.*) étoit autrefois en Angleterre un nombre ou une compagnie de dix hommes avec leurs familles, formant ensemble une espece de société, & qui tous étoient obligés de répondre au roi de la conduite tranquille les uns des autres.

Il y avoit dans chacune de ces compagnies un principal chef qui étoit appelé *dixenier*, du nom de son office ; & encore à présent dans quelques contrées ce mot est en usage, quoique cet officier ne soit maintenant autre chose qu'un commissaire, & que l'ancienne coutume des *décuries* soit tombée depuis longtems. *Chambers.* (G)

Ces sortes de dixeniers se sont conservés dans la police de la ville de Paris & de plusieurs autres villes de ce royaume, où l'on trouve des quarteniers pour chaque quartier, puis des cinquanteniers, quatre par chaque quartier, & des dixeniers qui sont ou doivent être seize dans chaque quartier. Autrefois ils avoient droit les uns & les autres d'assembler les bourgeois de leurs départemens ; mais depuis l'établissement d'un lieutenant général de police, ces offices de ville sont des titres sans fonctions. (a)

Une des premières et des plus légitimes questions que se pose un lecteur moderne est celle de l'auteur, et pour peu qu'il sache (et que l'édition lui propose l'accès facile à la page de titre) que cette *Encyclopédie* de Diderot et D'Alembert est le fruit du travail de nombreux collaborateurs, la seconde est : par quel moyen le reconnaître ? Pour que l'édition numérique réponde à ces interrogations, avant toute autre identification, les deux signatures « (G) » et « (a) » que comporte le texte de l'article doivent être caractérisées comme telles¹³, et donc spécifiées au moyen de métadonnées. Les lecteurs avertis savent que le (G) désigne l'abbé Mallet et le (a) Lenglet Du Fresnoy : mais de quelle partie du texte sont-ils respectivement les auteurs et pourquoi ? La réponse à ces questions se

13. Ce n'est pas le cas de la mention « *Chambers.* », dont on sait qu'elle désigne l'emprunt de tout ou d'une partie du texte à la traduction de la *Cyclopaedia*.

trouve, par bribes, dans les textes d'escorte de l'*Encyclopédie*, à des endroits connus des spécialistes. Ainsi, certaines correspondances entre signatures et auteurs sont faites à la fin de l'Avertissement du tome I, d'autres dans le tome III¹⁴, etc. L'éditeur Diderot se reconnaît parfois par l'*absence* de marque, parfois par une étoile *au début* de l'article ; et les encyclopédistes indiquent que si « plusieurs articles appartenant à la même matière », « immédiatement consécutifs », sont « faits ou revûs » par « la même personne », la désignation n'apparaît qu'« à la fin du dernier de ces articles »¹⁵. Ainsi, la simple attribution des articles signés, qui repose sur le décryptage des signatures et la détermination des pans de texte auxquels elles sont relatives, se décline en une multitude de cas particuliers, qui se présentent différemment d'un bout à l'autre d'une entreprise qui exigea plus de vingt années de travail, et dont nous ne dressons pas la liste ici¹⁶. Cela justifie par l'exemple le besoin, pour le lecteur novice comme pour le lecteur averti, que ce qu'il *voit* dans le texte – un (G), un paragraphe – soit *lié* à l'information *critique* (la marque est celle de l'auteur Mallet, et ceci est un paragraphe attribué à Mallet...), et que cette *liaison* fasse elle-même l'objet de justifications précises et développées qui s'appuient sur le texte (... émanant de la liste des auteurs figurant dans l'Avertissement du tome I, ou bien parce que, toujours suivant le même avertissement, l'article suivant est aussi un article d'histoire ancienne, par exemple).

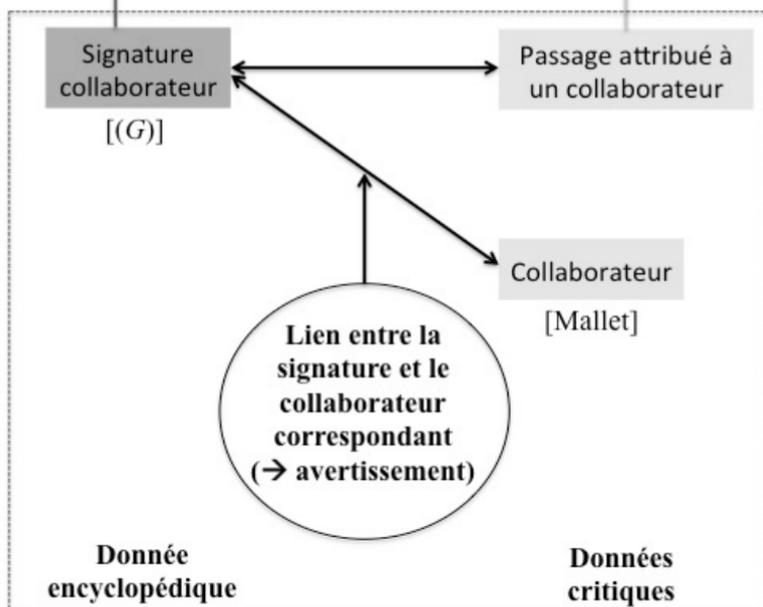
On voit ici se dégager un processus de structuration générique entre, d'un côté, des données qui appartiennent au découpage du texte, que nous appellerons des *données encyclopédiques*, et, de l'autre, des données qui supposent une interprétation et demandent raison, et que nous appellerons des *données critiques* :

14. Pour l'identification des signatures (G) et (a), voir *Encyclopédie*, t. I, p. xlvj et t. III, p. 905, respectivement.

15. *Encyclopédie*, t. I, p. lxxvj.

16. Pour de nombreux exemples de « polyphonies » à l'œuvre dans l'écriture encyclopédique, voir Alain Cernuschi, *Penser la musique dans l'Encyclopédie*, Champion, Paris, 2000, et plus particulièrement p. 661-663 pour les signatures.

DECENNA ou DÉCURIE, (*Hist. anc.*) étoit autrefois en Angleterre un nombre ou une compagnie de dix hommes avec leurs familles, [...] l'ancienne coûture des *décuries* soit tombée depuis longtems. *Chambers.* (G)



Entre une donnée encyclopédique (la signature présente dans le texte) et deux données critiques (l'auteur correspondant à la signature et l'extrait du texte de l'article à laquelle cette signature correspond), des liens sont établis pour rendre compte de l'attribution du passage à son auteur, *liens* qui *doivent pouvoir être annotés*, par exemple en renvoyant à la partie de l'Avertissement qui explique que la signature (G) désigne Mallet (mieux même, en ouvrant une fenêtre qui affiche cette partie) : « Mallet » est une donnée critique, qui peut être dotée de toutes sortes d'attributs, par exemple d'une notice biobibliographique.

Une telle structure, simple dans son principe, permet d'intégrer la complexité des cas particuliers, depuis l'erreur typogra-

phique¹⁷, jusqu'à l'absence¹⁸ ou à l'erreur d'attribution¹⁹. À partir de là, de multiples réseaux de circulation et de nouvelles possibilités d'interrogation du corpus s'ouvrent à nous : par exemple, collecter et rechercher dans tous les passages liés à une étoile, ou, plus finement encore, dans tous les passages attribués à Diderot parce qu'une signature « * » les précède, ou parce qu'un chercheur dont les travaux font autorité les a attribués à Diderot, ou encore, comparer ces listes.

Cette conception novatrice du *lien informé (ou annoté)* peut aussi être étendue aux planches, qui ont été négligées par l'édition numérique, sans doute en raison des difficultés rencontrées à interroger ces volumes en mode plein texte ou à relier telle planche aux articles qui y renvoient. La question est complexe, car ne relevant que fort peu de l'automatisme, compte tenu des modifications liées aux délais de publication des différents volumes et surtout au changement intervenu dans la conception même de la Description des Arts²⁰. Néanmoins, la collaboration des spécialistes du sujet permet d'envisager des solutions, sachant en outre que les grandes pages in folio des planches peuvent être subdivisées à l'instar des « passages » de texte.

En précisant clairement en amont ce qui est d'ordre textuel et ce qui est d'ordre interprétatif, ainsi que les relations que ces métadonnées entretiennent entre elles, le travail de structuration de données garantit en outre les conditions de réalisabilité d'une

17. « On a mis (Z) pour (Q) à l'article *Aide de Camp* » (*Encyclopédie*, t. I, p. xlvj).

18. Voir par exemple les arguments mobilisés par J. Lough, ouvr. cité, dans le chapitre « D'Holbach's contribution ».

19. Les errata du tome V signalent, par exemple, une erreur d'attribution : « Consultation, (*Méd.*) pag. 109; col. 1, lig. 39. Article de M. Bouillet fils, *lis.* de M. d'Aumont. » Un autre cas est celui des découvertes récentes, qui, elles, relèvent d'un apport critique. Ainsi, l'entrée ALLÉES DE JARDIN a une première partie terminée par (K), suivie d'une partie commençant par *, et donc attribuée à Diderot jusqu'à ce que Fabrice Ferlin découvre que non seulement le contenu, mais une remarque explicite de D'Alembert dans ses *Opuscules mathématiques* (t. III, 1764), permet de lui attribuer sans aucun doute la rédaction de cette partie de l'article (F. Ferlin, « D'Alembert et l'optique », *Recherches sur Diderot et sur l'Encyclopédie*, 43, 2008, p. 127-144).

20. J. Proust, ouvr. cité, p. 207 et différents travaux de Madeleine Pinault Sorensen.

interface de lecture et de navigation permettant d'éviter au lecteur toute confusion entre ce qui relève respectivement de l'*Encyclopédie* originale et des recherches dont elle a fait l'objet, et par là même, de l'informer *en toute connaissance de cause*. Cette étape, réalisée grâce à une étroite collaboration entre chercheurs et ingénieurs, pose en retour des questions informatiques substantielles. En effet, le modèle informatique qui assure la tenue du projet éditorial présenté plus haut est un modèle complexe de structuration et d'exploitation des données dans lequel les liens (hypertextuels par exemple) sont davantage que de simples associations entre deux éléments informationnels : ils sont à proprement parler porteurs d'informations, dont l'explicitation, l'explication ou la justification s'incarnent, éditorialement, sous forme de note.

Dans la continuité d'une réflexion sur la structuration des données se pose bien sûr aussi la question de l'interfaçage, ou plus précisément des diverses interfaces au moyen desquelles la chaîne éditoriale numérique complète est assurée. Il s'agit d'une part de l'interface d'édition, qui devra concrètement permettre aux chercheurs sur l'*Encyclopédie* de toutes origines disciplinaires de réaliser efficacement leur travail d'éditeur. Il s'agit d'autre part de l'interface de lecture, de navigation et de recherche consultable par les internautes. Il s'agit enfin, au confluent des deux, d'offrir un certain nombre d'outils de recherche et d'analyse utilisables par les lecteurs comme par les éditeurs du projet.

Du point de vue de la première exigence, l'interface collaborative d'édition de l'*Encyclopédie* disponible sur Wikisource a beaucoup à nous apprendre. L'expérience montre en effet à quel point le travail de collationnement et d'enrichissement collaboratif a pu être efficace : à partir de versions numériques existantes, et grâce à des procédés d'ocrisation évolués et mis à la disposition de tous, la plus grande partie de la transcription a pu être vérifiée en l'espace de quelques années, les formules mathématiques ajoutées, les tableaux et tables édités, les renvois entre articles précisément renseignés, etc. S'il ne nous semble pas souhaitable, pour des raisons scientifiques évidentes, d'ouvrir l'interface d'édition critique à tous les lecteurs, il paraît en revanche nécessaire d'opter pour un mode de travail collaboratif similaire entre les futurs éditeurs du projet. Nous pouvons ainsi

imaginer une plateforme partagée, dédiée au collationnement et à la réalisation de l'appareil critique, et par là même un processus d'édition critique dynamique, ancré dans le long terme, où viendraient se croiser apports anciens et nouveaux, et se conjuguer des compétences disciplinaires multiples et complémentaires. Dans notre cas, développer les outils informatiques permettant d'envisager ce type d'interface est en effet la condition *sine qua non* à la réalisation d'une entreprise d'aussi grande envergure, qui porte, rappelons-le, sur plus de 70 000 articles et plus de 2000 planches.

Dans cette perspective, un groupe de chercheurs aux compétences complémentaires devient bien sûr primordiale. Soixante ans de recherches depuis les travaux fondateurs de J. Proust et de J. Lough, puis l'*Inventory* de Schwab ont cependant permis « d'entrer dans la forteresse²¹ » en évitant les douves du contresens, par de multiples approches méthodologiques ; une équipe internationale d'une trentaine de chercheurs²² est donc déjà constituée, restant bien sûr ouverte à tous les spécialistes désireux de la rejoindre. Elle s'appuie sur une revue, *Recherches sur Diderot et sur l'Encyclopédie*, fondée en 1986, et sur le Groupe d'édition critique des *Ceuvres complètes* de D'Alembert, soutenu par le Comité D'Alembert de l'Académie des sciences.

Reste néanmoins à concevoir une chaîne de validation éditoriale pensée en conséquence, ainsi qu'une interface adaptée aux pratiques de travail très diverses d'une équipe d'éditeurs dont il faut présupposer, afin de n'en exclure personne, qu'elle ne possède pas de compétence informatique particulière (si ce n'est la maîtrise des logiciels de traitement de textes les plus courants). Pour réussir, une étroite interaction entre les informaticiens et ingénieurs qui vont concevoir

21. Piste de travail essentielle lancée par J. Proust, « Questions sur l'*Encyclopédie* », *RHLE*, janvier-février 1972, p. 52.

22. Parmi lesquels l'équipe du professeur Sumi de l'Université Keio de Tokyo et d'autres universités japonaises, largement engagée depuis plusieurs années dans le travail de repérage des métadonnées de l'*Encyclopédie*, ainsi que de nombreux chercheurs appartenant à divers centres de recherches scientifiques : l'Institut de mathématiques de Jussieu (UPMC), les universités de Lausanne, de Paris-Ouest-Nanterre, le laboratoire LIRE (Université Lyon 2), l'IRCL (Université Montpellier 3), le SYRTE (Observatoire de Paris), l'Institut universitaire européen de Florence, etc. Le comité de pilotage en France est constitué de Vincent Barrellon, Alexandre Guilbaud, Olivier Ferret, Marie Leca-Tsiomis, Irène Passeron.

l'interface, et les chercheurs amenés à l'utiliser paraît indispensable : les premiers proposeront, les second testeront, et ainsi de suite.

Ce dialogue constant s'applique bien sûr aux deux autres versants de l'interface : celui de l'édition accessible et interrogeable en ligne, et celui des outils d'analyse et de recherches proposés aux lecteurs et aux éditeurs. Là encore, des tests par un public de tous horizons (amateur et spécialisé) seront nécessaires, afin de parvenir à une interface de navigation et de recherche à la fois intuitive, ergonomique, et strictement compatible avec le principe de séparation claire entre ce qui relève de l'œuvre originale et de l'apport critique (selon la structure de données préétablie).

Du point de vue informatique, la conception de ces différentes interfaces à partir de la structuration des données constituée en amont pose plusieurs défis. Au delà des nécessaires interactions avec les chercheurs en sciences humaines et sociales au cours des phases de développement des outils se pose aussi la question de leur généricité. La perspective est double du point de vue éditorial : il s'agit à la fois de construire une plateforme capable de s'adapter à des modifications substantielles de la structure de données établie (il n'est pas à exclure que cette structure puisse en effet être ultérieurement retouchée en fonction des dernières avancées des recherches sur l'*Encyclopédie*), et capable de prendre en charge d'autres types de corpus répondant à des problématiques de recherche et éditoriales suffisamment similaires. C'est dans cette optique que le projet ENCCRE travaille actuellement en collaboration avec des informaticiens de l'INSA (LIRIS) et d'autres projets d'édition numérique (Archives Desanti, Projet Stendhal, Projet des « Dossiers de Bouvard et Pécuchet de Flaubert) autour d'« une approche générique pour la construction d'éditions électroniques de corpus documentaires multistructurés ».

En lien étroit avec ce travail informatique fondamental, de nombreuses autres « briques » seront développées par des ingénieurs, régulièrement soumises aux tests de l'équipe de chercheurs-éditeurs, et finalement assemblées pour former l'édition finale. L'ensemble sera conçu et développé de façon suffisamment générique pour être utilisable pour d'autres encyclopédies ou dictionnaires du 18^e siècle, ou d'autres œuvres possédant le même niveau de complexité éditoriale. Il permettra, *via* une interface publique,

une lecture et une navigation conformes à la politique éditoriale, et *via* une interface qui possèdera plusieurs niveaux d'accès, la continuation de l'édition critique collaborative. Des comptes utilisateurs permettront au lecteur de faire ses propres sélections de passages, suivant des critères explicites, de faire des comparaisons à bon escient par des outils adaptés, de profiter de l'apparat critique accumulé et constamment réinvesti et complété par la recherche.

Pour conclure, si l'éditeur sur support numérique a donc l'avantage, par rapport à l'éditeur papier, de pouvoir offrir au lecteur le choix entre plusieurs lectures, il a aussi pour contrainte de devoir réfléchir en amont à ces choix. Il ne peut faire l'économie d'aucune des obligations qui incombent à l'éditeur sur support papier (choix du texte, transcription fidèle, choix et justification des ajouts éditoriaux, présentation de l'apparat critique en fonction du lectorat visé), qu'il voit multipliées par la large variété des lectures virtuelles possibles. Trois formes de compétences et d'apports doivent donc interagir dans l'élaboration d'une telle édition numérique : celles du chercheur sur l'*Encyclopédie* (fondée sur les études existantes et sa propre analyse du contenu, du genre, de la forme d'écriture et d'argumentation), celles du chercheur en informatique (fondée sur la compréhension des attentes du premier et sa connaissance de l'état de l'art, en termes de structures et de problématiques informatiques), et celles de l'ingénieur (fondées sur la connaissance des outils d'encodage et leurs développements conceptuels, spécifiques, ainsi que sur la capacité à développer des interfaces ergonomiques). Aucune priorité ici, mais un dialogue, nécessairement complexe, nécessairement fécond.

Alexandre GUILBAUD
UPMC, Institut de mathématiques de Jussieu

Irène PASSERON
CNRS, Institut de mathématiques de Jussieu

Vincent BARRELLON

Olivier FERRET
(Université Lyon 2, laboratoire LIRE)